

The Prime Machine HD Guide

I want to ... learn more about collocation

If you are an undergraduate or postgraduate student, using English as the language of instruction, you will need to develop your academic writing skills. Teachers and textbooks will often encourage you to pay attention to combinations of words which make the expression of ideas concise and clear. When we look in a corpus at the combinations of words which occur, we can see that words have a tendency to co-occur with a smaller set of other words; the computer can determine the strength of these relationships by calculating the likelihood of two or more words co-occurring in relation to their frequencies in the corpus overall. You can select a general corpus containing many different kinds of texts (such as the complete British National Corpus), or a corpus of academic texts from a variety of disciplines (such as the BNC: Academic), or a corpus of academic texts from a specialized field (such as the Hindawi academic corpora, or a corpus you build yourself). The results of these statistical processes mirror the relationships between words which must exist in the minds of the communities of speakers and writers as they share, discuss and create knowledge in their specialist areas. The results can help you notice combinations of words which are likely to seem more natural and help you become more aware of combinations which may be considered more rare or creative. If you put words together just based on basic meanings from wordlists or dictionaries, you will probably discover that your EAP teacher and subject professors find that your expressions seem unnatural or awkward or simply just don't seem to fit together. When you are trying to communicate a new idea, or you want to use language more creatively, knowing the collocations which already exist will make your new combinations stand out. This guide will explain the steps you can take to deepen your knowledge of collocations through game activities and analysis using The Prime Machine HD corpus tool.

Steps to complete:

1. Learn how to spot collocations in concordance lines and other corpus data by playing some games about academic synonyms and prepositions.
2. Learn how to use the search support features to help you find combinations of words to compare.
3. Learn how you can build your own corpus, and view collocations of words of interest within it.
4. Explore the collocational strength between words in your own sentences, by using an online corpus as a baseline.

This approach can help you:

1. Learn how to notice the typical patterns of usage of different words and collocations in general English, academic English and in specialist fields.
2. Find combinations of words from your own writing which may need reformulation.
3. Discover the patterns of usage of collocations which may be distinct from the words from which the combinations are formed.

What you'll need to get started:

- The Prime Machine HD for Windows, macOS, iPad, iPhone or Android (available free from <https://www.theprimemachine.net/>)
- Some combinations of words to inspire your exploration after you've tried the games - perhaps phrases you want to learn more about or some sentences from your own writing to check for collocational strength.
- Patience, enthusiasm and an open mind!

Looking at your use of language through linguistic spectacles

You don't need to be a student of linguistics to be able to analyse the patterns of use of language (but of course if you've studied linguistics you can draw on your linguistic knowledge). When you complete an academic assignment, you are essentially showing to your professors, the institution and the world that you are becoming a member of an academic community – a community which communicates using academic conventions and through academic language. Your use of academic vocabulary and specialist terminology reveals your experience reading and hearing these words and phrases as they are applied in your specialist discipline. Some collocations can be found in different disciplines and across text types, but others can have special meaning for members of a specialist field of study.

Are you ready to become a language detective?

To make the most of the approach described here, you will need to try to think about your word choices as a detective might look at clues. As well as being an amazing expert in detection, Sherlock Holmes is also remembered in the stories as being an expert in disguise.

- As a detective, can you look at a hundred examples of a word or phrase and pick out some of the patterns which occur?
- As a detective, can you look at the context and co-text of examples from a corpus to see deeper meanings and typical uses of words and collocations?
- As an expert in disguise, can you dress-up your own writing, so as to blend into the norms and expectations of expert language users?

Getting started

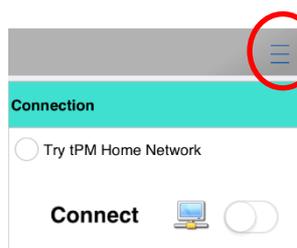
The best place to get The Prime Machine HD (tPM) is from an official store. It is free!



Windows and Android users can also download the App directly from the website:

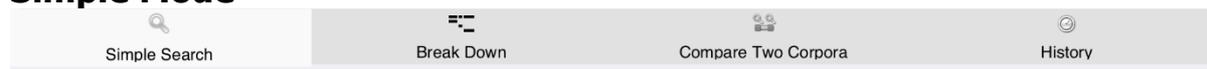
<https://www.theprimemachine.net>

When you first use tPM, you almost certainly will want to connect to the server to access pre-prepared corpora and resources. There are three modes: Simple Mode, Full Mode and Game Mode. The Full Mode includes additional tabs and features for corpus research and DIY corpus work.



The main 'hamburger' or 蒸笼 menu in the top-right corner allows you to connect and change mode.

Simple Mode



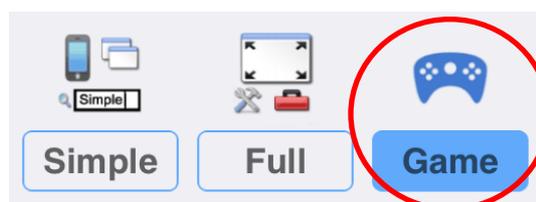
Full Mode



Game Mode

Game Mode provides a different user interface, giving you a list of groups of words to study before you *take a challenge* and try to guess which word matches a set of concordance lines.

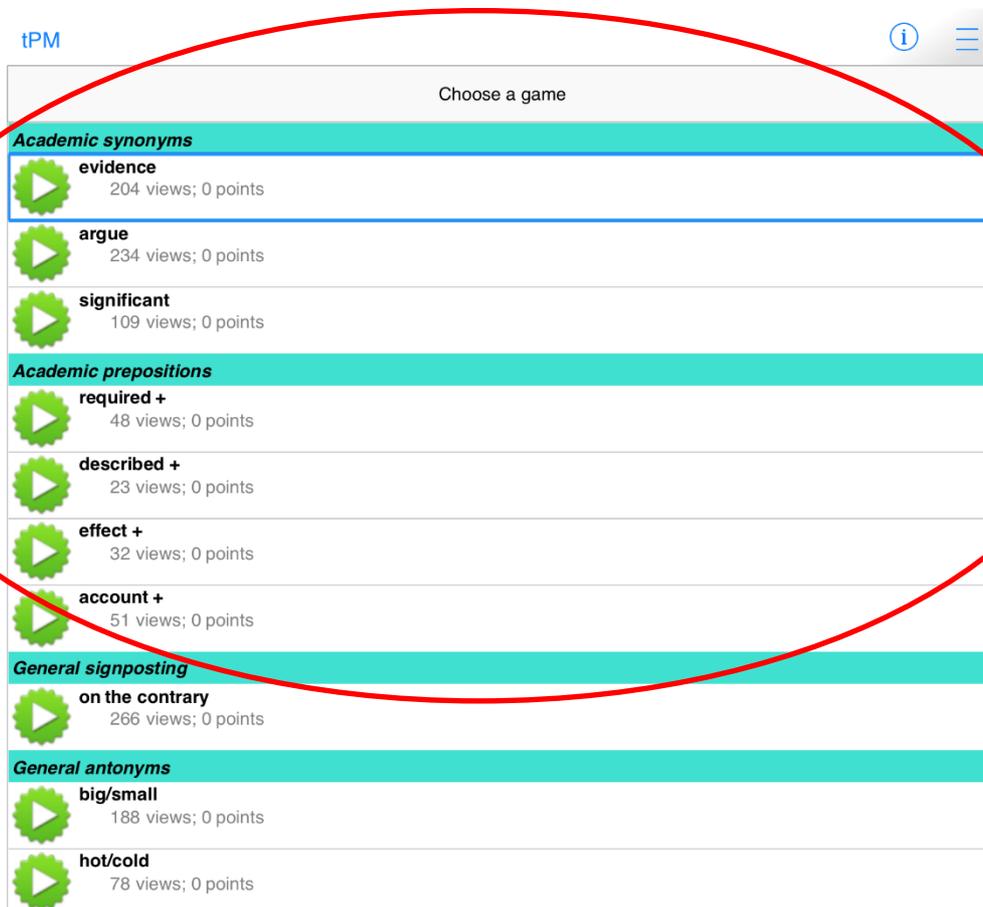
Two groups - **Academic synonyms** and **Academic prepositions** - are good places to start if you want to learn how collocations can help you choose between synonyms and find differences in the meaning of verbs when followed by different prepositions.



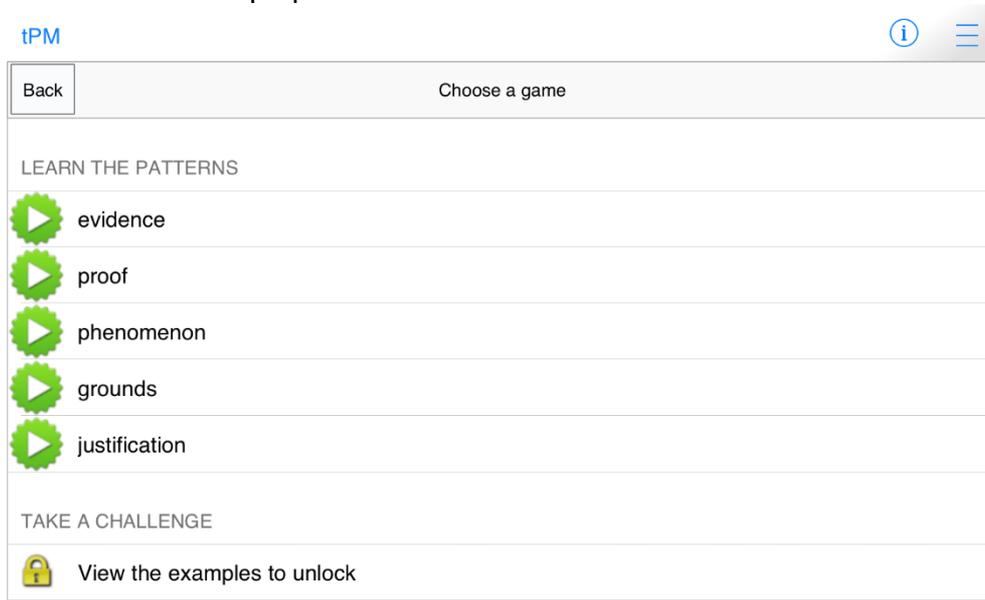
Step 1: Learn how to read corpus data using tPM Game Mode

How to play

1. Choose Game Mode before connecting to the tPM Server. Then Connect.
2. The first time you try it, you will be shown the instructions for the game.
3. Choose a set of words from one of the games listed. In this worksheet, we are interested in academic synonyms and academic prepositions, so select one of the words from the first or second group.



4. When you have selected a game, you will see a submenu containing the words or combinations you have to learn. Open the items in the group one by one, looking at the concordance lines and thinking about the patterns you see. This is your chance to try to notice and remember the patterns of usage for each of the different words and combinations of verbs and prepositions.



5. When you select each word, 200 examples will be downloaded from the server. The displays for learning the words are the same as when you look the words up in Simple Mode.

Collocations

BNC: Academic: **evidence** 200

	to the left of node	Node	Text to the right of
1	ve said there is no	evidence	that such drastic ar
2	that there was no	evidence	that the combinatio
3	tions, there is little	evidence	that sheep in ender
4	nths, there was no	evidence	to suggest that ren
5	There is also some	evidence	that cities doing we
6	d/ There is some	evidence	that he received fo
7	There is also some	evidence	that couples are liv
8	in fact there is no	evidence	of overlaps or rival
9	is no independent	evidence	for this, and even if
10	here is no reliable	evidence	for 'the present da
11	there is no certain	evidence	that Ecgberht was
12	prepared to give	evidence	that he experieced
13	, but there is now	evidence	that it can safely be
14	there is sufficient	evidence	to charge him (s.37
15	gery. There is still	evidence	that the young and
16	there any positive	evidence	that might prompt
17	olice and Criminal	Evidence	Act 1984, stemmed
18	that there is some	evidence	from two studies th
19	t... There is some	evidence	, though not very m
20	There is abundant	evidence	from a number of s

This menu button allows you to change the display settings for the current tab of results.

You can sort the lines in different ways to help you notice patterns.

This shows the number of examples already downloaded.

When learning words in Game Mode there are three kinds of data to view: Lines, Frequencies and Collocations.

On a wider display (like this iPad), you can see a concordance card showing more information about the currently selected line.

The screenshot shows the iPad interface with a table of concordance results. The table has four columns: 'Text to the left of node', 'Node', and 'Text to the right of node'. A red box highlights a specific row in the table. To the right of the table, a detailed concordance card is displayed, showing the selected text and its context. The card includes a title 'Politics, Law and Education' and a citation 'The Weekly Law Reports 1992 Volume 3. u.p.'. The card also contains a paragraph of text with the word 'evidence' highlighted in blue. At the bottom of the screen, there are three tabs: 'Lines', 'Frequencies', and 'Collocations'. The 'Lines' tab is currently selected.

Lines +

On narrower displays (like this Android phone) you can view either lines or cards. On all devices, switch between Lines + (if space allows), Lines and Cards using these settings.

The screenshot shows the settings menu for the application on an Android phone. The menu is titled 'Lines +'. It contains several options, each with a small icon and a checkmark indicating it is selected:

- Lines
- Cards
- Wider Context
- Sentence Only
- Show Citations
- Hide Citations

Other useful settings include just showing one sentence for each hit and hiding the citations.

	Text to the left of node	Node	Text to the right of node
1	on 61 of the Taxes Management Act 1980, but as I have said there is no	evidence	that such drastic and highly unusual proceedings were either threatened
2	Parker LJ rejected that submission by pointing out that there was no	evidence	that the combination of the two ingredients was commonly known or used
3	Second, in contrast to many other helminth infections, there is little	evidence	that sheep in endemic areas develop an effective acquired immunity to H
4	Although mean follow up was only 49.9 months, there was no	evidence	to suggest that removal of stones resulted in any improvement in prognos
5	There is also some	evidence	that cities doing well in manufacturing will also do well in service-sector er
6	/ There is some	evidence	that he received formal academic training at Cambridge.
7	There is also some	evidence	that couples are living together in a stable relationship without being marr
8	relationship sounds, in fact there is no	evidence	of overlaps or rivalries.
9	But there is no independent	evidence	for this, and even if the albedo in the past was zero, any change would pr
10	concludes that there is no reliable	evidence	for 'the present day vulgarism'before the eighteenth century (p. 296).
11	ians, after 765 there is no certain	evidence	that Ecgberht was for long, if at all , 'a mere dependant' of Offa./
12	was the victim is prepared to give	evidence	that he experienced harassment, alarm or distress, so much the better.
13	s after overdose, but there is now	evidence	that it can safely be given to patients up to 24 hours after ingestion and pe
14	ermine whether there is sufficient	evidence	to charge him (s.37(1) and Code Para. 17).
15	There is still	evidence	that the young and newly diagnosed patients remain at increased risk of c
16	Is there any positive	evidence	that might prompt us to adopt the more complex position that differentiat
17	courts in which indictable offences are tried, and the Police and Criminal	Evidence	Act 1984, stemmed from Royal Commissions or similar forms of independ
18	/ It should be noted that there is some	evidence	from two studies that mean levels of serum cholesterol are higher in subje
19	There is some	evidence	, though not very much, of occasions when the plaintiff or husband acted c
20	There is abundant	evidence	from a number of studies which shows that. apart from health visitors. the

When Sentence Only is selected, you can see if the word occurs near the beginning or end of sentences easily.

Lines

This shows the kind of text from which this example comes with reference information below.

At the top of the card you can see strong collocations near the word of interest.

Politics, Law and Education
The Weekly Law Reports 1992 Volume 3. u.p.

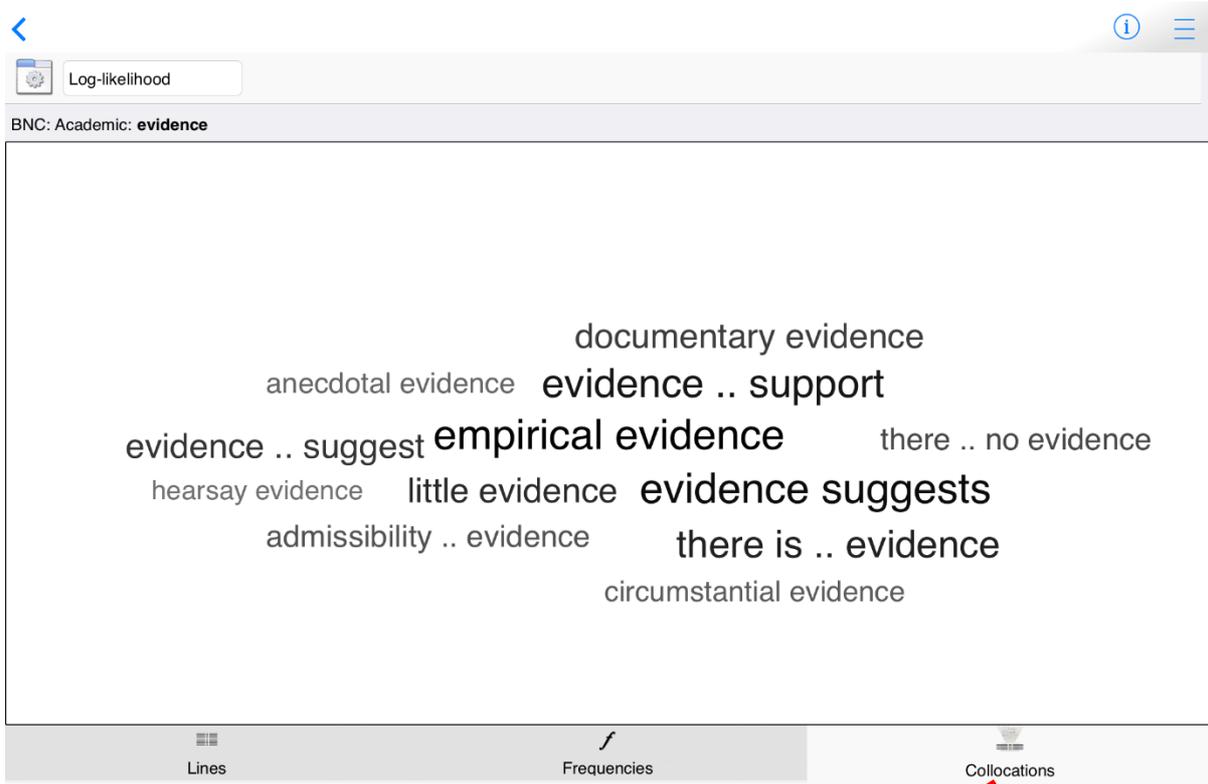
— cannot regard it as involving duress on the part of the revenue The position might be different if Woolwich had paid under threat of the revenue taking distress proceedings without a court order under section 61 of the Taxes Management Act 1980, but as I have said there is no **evidence** that such drastic and highly unusual proceedings were either threatened by the revenue or anticipated by Woolwich, still less that Woolwich had a reasonable apprehension of being put out of business by them.

Nolan J.'s reference to 'duress on the part of the revenue' is important —

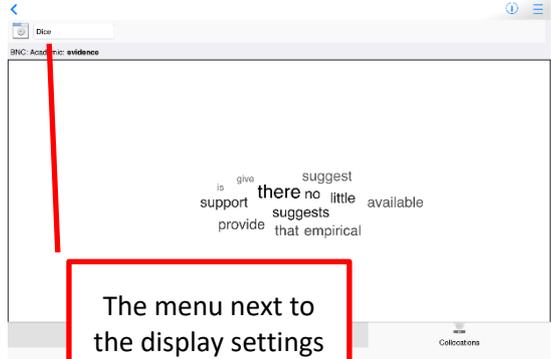
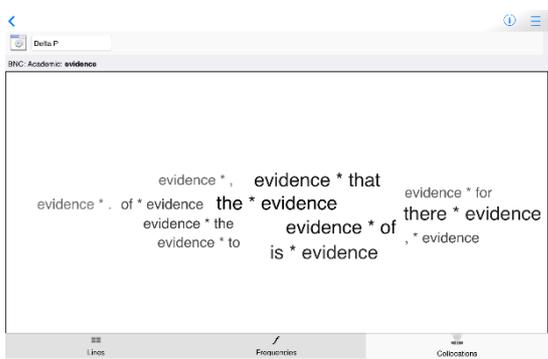
You can navigate through the examples using the sidebar.

The card shows up to one sentence before and after the sentence containing the hit. Text is displayed with paragraphing.

Cards



The Collocations Tab shows the word partnerships – the combinations of words which go together naturally in the corpus.

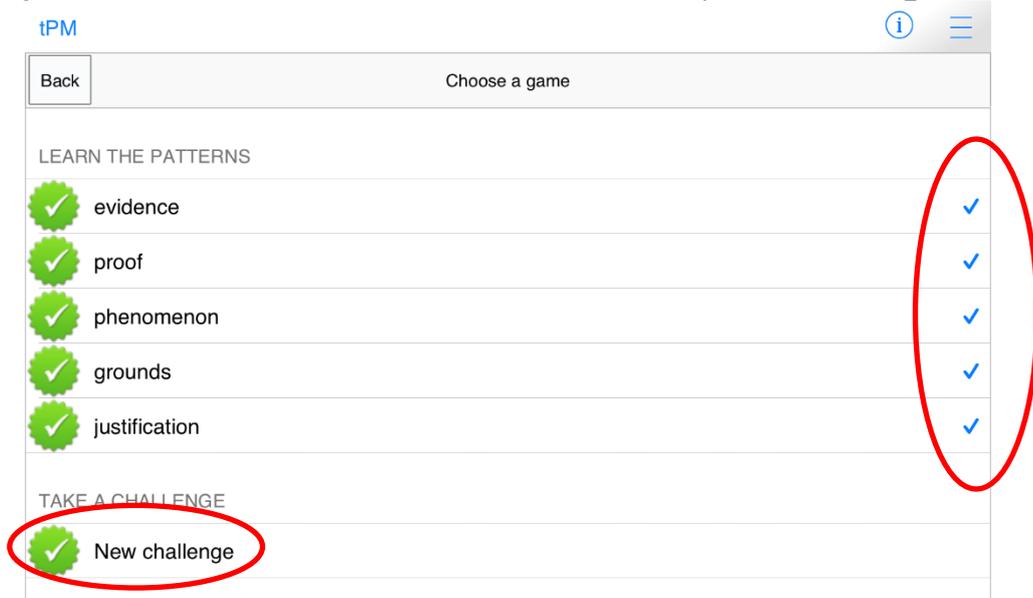


The menu next to the display settings button allows you to choose different statistics.

Collocate	Frequency	L4	L3	L2	L1	R1	R2	R3	R4	T-Score	MI	Dice
1 yes	1,416	16%	14%	21%	-	1%	8%	2%	1%	31.74	27.03	0.08
2 suggests	194	1%	1%	1%	-	17%	13%	6%	7%	13.64	22.50	0.03
3 no	642	3%	2%	21%	-	1%	7%	1%	-	24.63	24.70	0.08
4 little	229	1%	5%	16%	73%	-	-	1%	-	14.23	21.45	0.03
5 support	193	2%	2%	2%	-	1%	72%	11%	11%	13.69	21.12	0.03
6 provide	164	4%	10%	40%	43%	-	1%	1%	2%	12.90	20.62	0.02
7 that	2,062	3%	3%	6%	2%	48%	13%	13%	7%	44.74	26.58	0.08
8 empirical	123	-	1%	16%	-	-	1%	-	-	19.14	21.12	0.02
9 available	145	1%	1%	6%	41%	23%	11%	10%	3%	11.63	20.25	0.02
10 suggest	118	1%	1%	-	-	6%	64%	6%	3%	10.74	20.14	0.02
11 give	125	3%	5%	16%	72%	-	1%	2%	3%	10.96	19.51	0.02
12 is	2,921	6%	13%	36%	16%	3%	6%	6%	-	42.90	26.10	0.08
13 some	332	6%	8%	10%	61%	-	8%	3%	4%	17.43	21.51	0.02
14 show	106	3%	5%	11%	25%	1%	48%	1%	6%	10.11	19.23	0.02
15 further	143	3%	4%	7%	81%	-	3%	1%	1%	11.81	19.36	0.02
16 critical	66	-	6%	7%	72%	-	8%	5%	1%	6.14	16.64	0.01
17 case	589	6%	6%	12%	31%	11%	7%	7%	-	22.08	22.53	0.07
18 sufficient	81	2%	-	4%	61%	-	4%	9%	-	6.69	16.98	0.02
19 any	296	4%	5%	14%	53%	1%	24%	8%	7%	14.60	20.15	0.01
20 process	78	4%	8%	16%	31%	3%	1%	3%	-	8.57	18.51	0.01

Different statistical measures can be used to create lists of collocations. You can also view lists of collocations as tables.

6. When you have viewed concordance lines for all the words, **New challenge** will unlock.



7. Each time you take a new challenge, decide how confident you are that you will be able to guess each word correctly.

The screenshot shows the confidence selection screen. At the top, there is a back arrow and a question: 'How many lines do you think you need?'. Below this is a slider with markers at 25, 50, 75, and 100. The slider is currently set to 50. Below the slider is another question: 'How confident are you?'. There are five rows of words: 'evidence', 'proof', 'phenomenon', 'grounds', and 'justification'. Each row has a blue progress bar and a '3/5' label. A 'Confirm' button is at the bottom. Three red callout boxes provide instructions:

- Top callout: 'The more lines you have, the easier it will be to guess, but the fewer games you can play overall.'
- Middle callout: 'If you slide the confidence up to 5/5, you will get more points for a correct answer, but face a heavier penalty if this is the missing word and you guess incorrectly.'
- Bottom callout: 'If you slide the confidence down to 1/5, you will get fewer points for a correct answer when this item is the correct answer, but you won't lose so many points when you guess incorrectly.'

Tips:

- Synonyms often have different collocations; if you learn the strong collocations, you should see some of these in the lines at the top (when the lines are sorted using Collocations – the default).
- Synonyms often have different grammatical patterns. Although they may all be the same word class, some nouns may be used with *the* or *this*, while others may be used with *a*, *an* or *some*. You may also find differences in use of modal verbs or prepositions, for example.
- For the academic prepositions games, it is better to try to explore the possible meanings and patterns of each combination of verb and preposition – while prepositions may have some general meaning, we often find that it is hard to generalize.
- If you find the challenge too hard, go back and read more examples of the words by selecting them from the **Learn** section. You can continue with the same challenge again later.

	Text to the left of node	Node	Text to the right of node
1	the OECD, claimed that 'up to now there is no firm	---	that the current high levels of unemployment are du
2	er insertion. However, in other areas there is good	---	that archaic and illogical ageist practices survive. Or
3	ghts (R C L Page et al, unpublished data) Thus no	---	to support this alternative hypothesis is available. Th
4	onsidered risk factors for eating disorders. There is	---	that anorexia nervosa itself has developed a positive

In landscape orientation, more co-text will be visible

	t to the left of node	Node	Text to the right of i
1	now there is no firm	---	that the current high
2	areas there is good	---	that archaic and illog
3	shed data) Thus no	---	to support this alterr
4	disorders. There is	---	that anorexia nervos
5	Street. There was	---	that passers-by wer
6	s far more empirical	---	about women's close
7	creative area but no	---	of neoplasia. Postop
8	aving documentary	---	to support particular
9	riods. This is further	---	of a somewhat depre
10	or example, there is	---	to suggest that a go
11	en he came to give	---	. If ever there was a c
12	ndustries. Since the	---	for their existence is
13	rovide the essential	---	for the rest of this se
14	ms to be fairly clear	---	in favour of an intera
15	From the available	---	, it seems fair to conc
16	the findings of fresh	---	;(d) consequent on th
17	ight turn out to be	---	against it. This exam]
18	spute about EModE	---	for overlap and merc
19	out very persuasive	---	indeed./ One of the t
20	3.2 Palaeomagnetic	---	/ During the mid-195

In portrait orientation, more lines will be visible and most collocations will still be in view.

	t to the left of node	Node	Text to the right of i
1	nces The burden of	---	of causation under tl
2	its business without	---	of special damage. M
3	ere is no conclusive	---	, it seems highly likely
4	f risk, the burden of	---	was reversed and th
5	case, the burden of	---	appears to be on the
6	ift in the burden of	---	: in defamation the de
7	ng of the burden of	---	for proving "reasonal
8	n on the burden of	---	: but the omission wa
9	bear the burden of	---	? Do the defence hav
10	ere is no burden of	---	placed on either part
11	d upon, the onus of	---	is firmly upon the de
12	everses the onus of	---	by placing it upon th
13	our being taken as	---	of the existence of ai
14	tion to expunge the	---	of a surety against th
15	ly sue for libel upon	---	of special damage; a
16	t a high standard of	---	is apposite in cases a
17	y a civil standard of	---	to a high degree of p
18	re any of these fool-	---	, and if not might I be
19	nally-based gens: a	---	of the Marxist view o
20	Lear is a sufficient	---	of the destructivene:

Look for the strong collocations you've learned to the left and right of the gap.

8. Look at the set of concordance lines and guess the missing word.

<
Collocations

50

	Text to the left of node	Node	Text to the right of node
1	87. (e) Defences The burden of	---	of causation under the CPA 1987
2	the way of its business without	---	of special damage.Morland J. hel
3).Whilst there is no conclusive	---	, it seems highly likely that this pie
4	n that area of risk, the burden of	---	was reversed and the defendant
5	e Schroeder case, the burden of	---	appears to be on the party arguir
6	are (1) the shift in the burden of	---	: in defamation the defendant has
7	and the placing of the burden of	---	for proving"reasonableness" whe
8	mit a direction on the burden of	---	: but the omission was not necess
9	the defence bear the burden of	---	?Do the defence have the resourc
10	est, where there is no burden of	---	placed on either party.The reason
11	fence is relied upon, the onus of	---	is firmly upon the defendant to pr
12	because it reverses the onus of	---	by placing it upon the seller (see
13	iolent behaviour being taken as	---	of the existence of an inner state
14	issued a petition to expunge the	---	of a surety against the estate of a
15	ority could only sue for libel upon	---	of special damage; and that the c
16	n accept that a high standard of	---	is apposite in cases affecting libe
17	ave to satisfy a civil standard of	---	to a high degree of probability tha
18	and why?Are any of these fool-	---	, and if not might I be wrong after
19	and communally-based gens: a	---	of the Marxist view of the State a
20	e end of King Lear is a sufficient	---	of the destructiveness of hypocris
21	not be too difficult to produce a	---	of evidence that can be used dire
22	n general should now be clear.	---	of 2.5.4 The cases p = 2, 3 are e;
23	formula#in Zn as required.The	---	of M1 is similar.Let us offer a stre
24	hat distinguishes this offence is	---	of an intention to kill, not the occu
25	n to the throne Sigismund gave	---	of his statesmanship when at his
26	false trade description to goods	---	of dishonesty is unnecessary for
27	obbes was amazed by Euclid's	---	of this complex idea.He was stun
28	in the text, about the difficulty of	---	: the case just discussed is a rela
29	t can be established.Moreover,	---	of non-consent on a charge of bu
30	, the most consistent method of	---	'(1983: 358).Religious texts also r

Politics, Law and Education
 Sale and supply of goods.
 Davies, Iwan Longman Group
 UK Ltd Harlow 1990 1-90

— account of any reduction in damages through contributory negligence Section 6(4) applies the Law Reform (Contributory Negligence) Act 1945 and s5 of the Fatal Accidents Act 1976 relating to contributory negligence, to claims under Pt I of the CPA 1987. (e) Defences The burden of ---of causation under the CPA 1987 is upon the victim. Damages may be reduced by his contributory negligence and, in this respect, it could be argued that disregard of instructions breaks the chain of causation. ...

Lines
Answer

When you are ready to guess, click or tap the Answer button.

Step 2: Using search screen support and formulating your own queries

After you have completed some of the games for learning about academic synonyms and academic prepositions, you should find that you have become more familiar with how to read and use concordance data. The next step is to make the most of the different ways to search for your own words and phrases.

There are four different search screens for you to use: Simple Search, Break Down, Compare Two Corpora and Multiple Searches (under Research Tools on Full Mode).

Simple Search

This is the main search screen and it allows you to enter one or two search queries at a time. If you enter a word or phrase and tap the "hamburger" or 蒸笼 menu inside the search box, suggestions will appear. You can see the number of examples in the currently selected corpus. You can also see collocations, other word forms and words with a similar meaning.

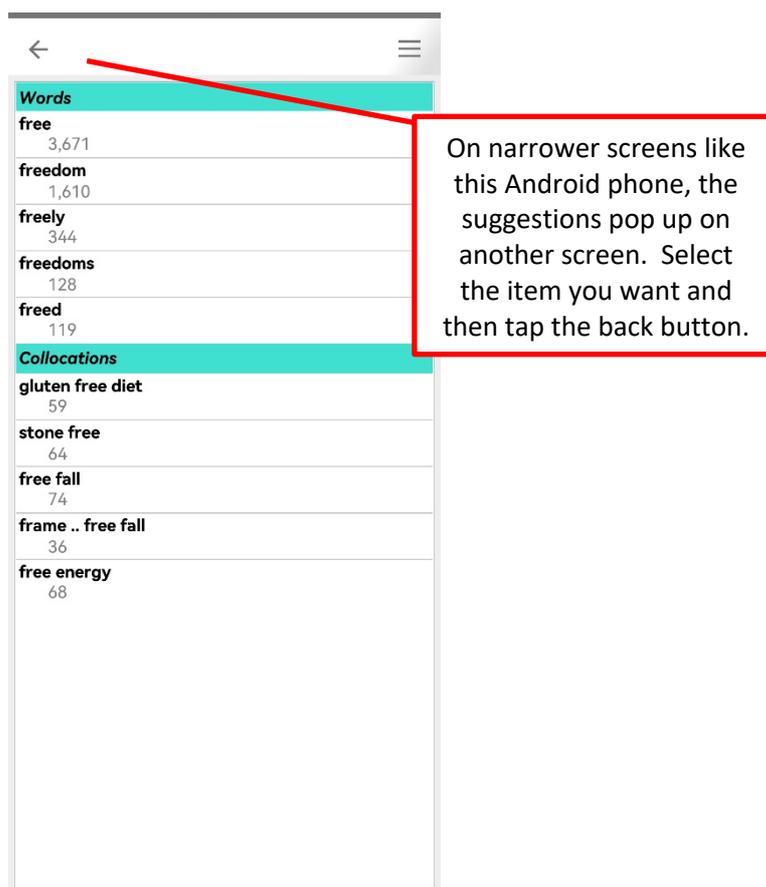
Frequency is important (and tells you the number of hits), but collocation strength is based on the statistical tests, trying to mirror the relationships between words of varying raw frequencies. The word *of* is very very frequent, but it will only show in a collocation if the statistical test shows its frequency in a combination with the search word is unexpectedly high.

The suggestions and the number of hits and the collocations will be different in different corpora. Click here to change to a different corpus.

This button will give you more information about the currently selected corpus.

The left-hand side shows single word suggestions and strong collocations. The numbers are the number of hits in this corpus.

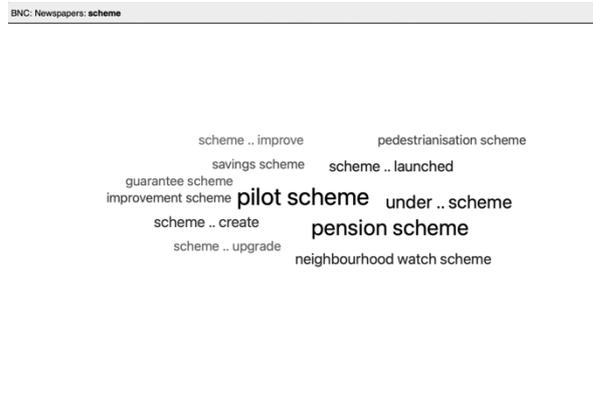
The right-hand side will show other word forms and words and collocations with similar meanings.



Tips:

- When you enter one word, the suggestions will be for longer strings with the same letters (auto-complete) and strong collocations on the left; and other word forms and words with similar meanings on the right.
- The suggestions for similar meanings use several different resources including English-English resources and words which have the same English translation in a Chinese-English dictionary. You can only search for English words, but some suggestions are based on a Chinese-English dictionary.
- If you click or tap the hamburger/蒸笼 button again, you will be given more suggestions including more collocations.
- When you enter more than one word, tPM will try to find a match in the list of strong collocations in this corpus. If no match is found, it does not mean the combination does not exist.
- From the Simple Search screen, you can force the app to show concordance lines and frequency data for combinations of words not stored as a LL collocation in the database, by using three special symbols:
 - Putting _ between words means these words must occur in order with no other words in-between.
 - Putting * between words means these words must occur in order, but there could be zero or more words in-between (up to a total phrase length of 5 words).
 - Putting | between words means these words can occur in any order but all must occur within a 5 word window.
- More advanced searches can be performed using the Multiple Searches tab.

When you are using a corpus tool, you will often need to perform several searches and to refine and retry different combinations. For example, if you look at the concordance line results for a noun you are interested in and you discover it has a strong collocation with a particular verb, you can return to the Simple Search tab and perform a new search for the two words in combination. Then up to 200 hits for the collocation will be displayed and you can explore the patterns of use for the collocation in more detail.



In BNC: Newspapers, the collocation cloud for *scheme* contains *scheme .. launched*.

This means that the verb *launched* has a strong statistical relationship with the noun *scheme*, taking into account the specific verb form and the word order.

Because there are over 2,000 hits for *scheme*, only a few of the concordance lines in the sample will contain the combination.

But if you return to the Simple Search screen and look up the collocation, all 20 lines will be retrieved.

The interest in the combination *scheme .. launched* may spark your interest in the verb *launched* – what other things are described as being *launched* in newspapers?

Returning to the Simple Search tab, you may want to explore the combinations of *launched* and also to explore the combinations of words with a similar meaning such as *established*, *published*, and *carried out*.

On the Break Down tab, we can enter a longer phrase or sentence and click the hamburger / 蒸笼 button inside the search box, and it will split the sentence into words, allowing us to look at combinations of up to five words starting at the currently selected item.

The first column shows the frequencies of each word in the sentence.

The second column shows combinations of words to the right of the selected word. If the combination is stored as a LL collocation it shows the number of hits.

If you click the + button, suggestions may appear.

You could check concordance lines if you couldn't guess "On the other hand" is better.

Choosing another word will show combinations with words to its right

These two words don't appear adjacent in this corpus, but they are nearby.

Perhaps *statistical techniques* would be better?

On the Compare two Corpora tab, you can enter one word or a phrase and then retrieve concordance lines and other corpus data from two different corpora in one operation.

Results from the main corpus will be displayed on the left (A).

BNC: Newspapers is a useful corpus for finding collocations for social issues or topics for general EAP essays.

Results from the second corpus will be displayed on the right (B).

These buttons allow you to change the second corpus or view information about it.

The suggestions will show the frequencies in both corpora, with collocation suggestions based on the strong patterns in the main corpus.

On the Multiple Searches tab (under Research Tools), you can perform a number of more advanced queries:

- Combining the results of more than one word form by entering multiple words separated by space, using wildcard pattern matching, or selecting the option to include word families;
- Forming patterns for phrases with multiple words in up to 5 slots, with options for the ordering of items and whether or not other words can occur in-between.
- Using filters to get results from one text category or advanced filters based on labels or features such as part-of-speech and position in text.

On the Multiple Searches tab as well as getting concordance lines and displaying them on the left (A) or right (B) you can also get frequency tables and plots.

tPM has been designed to provide easy access to concordance lines and other summary statistics for strong LL collocations. While you can look at other combinations using the special `_`, `*` and `|` symbols, or use the Multiple Searches tab to merge results for different word forms, you should be able to find many useful kinds of data by simply using the collocation searches on the Simple Search screen.

The screenshot shows the 'The Prime Machine HD V1.28' interface. At the top, there's a search bar with 'BNC: Academic' selected. Below it are various tool buttons like 'Simple Search', 'Break Down', 'Compare Two Corpora', 'Research Tools', 'DIY Tools', and 'Hist'. A main search area contains a 'Node word(s):' field with '+ crucial crucially' and a 'Co-text' section with four 'Co-text' fields and an 'Exclude' field. On the right, there's a 'Filters' panel with options for 'No filters', 'Single category' (set to 'Humanities and Arts'), and 'Advanced...'. Below that is an 'Other Settings' panel with checkboxes for 'Include word families', 'Use * wildcards', 'Ordered list', and 'No gaps'. At the bottom right, there's a 'Multiple Search Tools' panel with buttons for 'Get Frequencies', 'Get Plots', and 'Get Lines'. A red 'A/B' button is also visible. Five red callout boxes provide instructions: 1. 'Results will come from the main corpus; click or tap here to change.' (pointing to the corpus dropdown). 2. 'Some corpora are split into categories, so you can limit the search to specific groups of texts.' (pointing to the 'Single category' dropdown). 3. 'Enter lists of words separated by spaces to find any of these words and merge all the results.' (pointing to the 'Node word(s)' field). 4. 'If you enter one word and click or tap on the + button, other word forms for this word in the current corpus will be added to the box.' (pointing to the '+' button in the 'Node word(s)' field). 5. 'Clicking or tapping on the A/B button will send the results to the left (A) or right (B) so you can compare two sets of results side by side.' (pointing to the 'A/B' button).

The other settings panel allows you to specify:

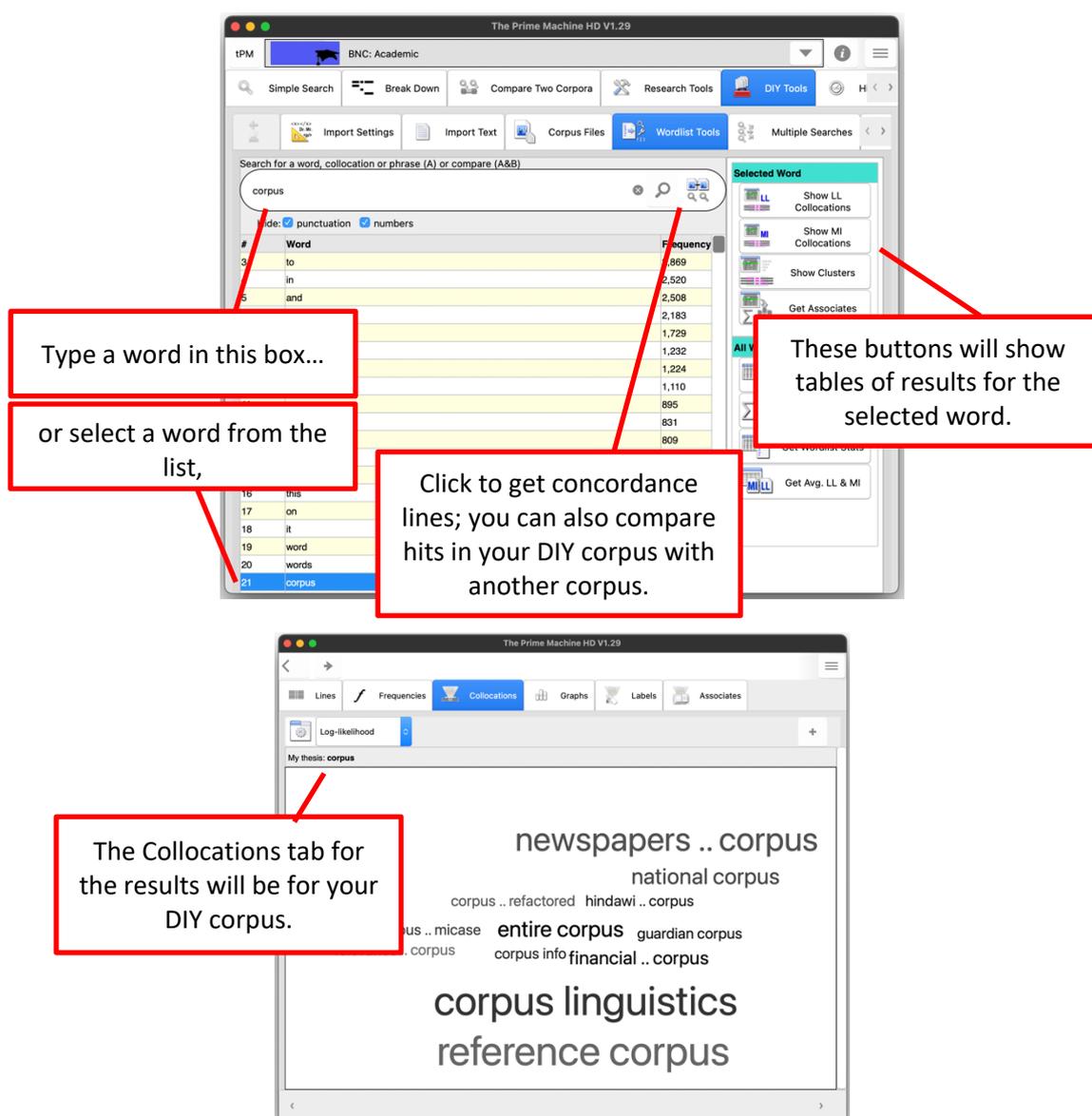
- Whether to include all the different word forms in the corpus for each word you have entered. For example, if you tick this option and enter the word *analysis* the results will include *analyses*, *analyse*, *analysing*, *analysed*, etc.
- Whether you want to use * to represent zero or more letters. For example, *analy** would include all words beginning with the letters *analy*.
- Whether the words entered in the co-text boxes need to occur in order or in any order and whether or not other words can occur in-between.

The 'Other Settings' panel is shown with the following options:
 Include word families
 Use * wildcards
 Ordered list
 No gaps

Step 3: Building your own corpus, and viewing collocations of words of interest within it

Building a corpus means gathering one or more texts and importing them into the corpus tool. For research, you would normally need to consider how you select your corpus texts and try to ensure your selection of texts is not made with unnecessary bias. But for a student or teacher wanting to explore collocations for a specific academic discipline, a good starting place would be to gather texts from the web pages, documents and articles used in a course. For step-by-step instructions on how to build your own corpus, see *The Prime Machine HD Guide I want to ... build my own corpus*. You can build a corpus on any platform (Windows, macOS, iPad, iPhone and Android).

Once you have your corpus ready, you can select a word from the DIY word list or type it in, and use the buttons on the right to show different kinds of collocations, or use the magnifying glass button to retrieve concordance lines and view collocations on the Collocations tab.

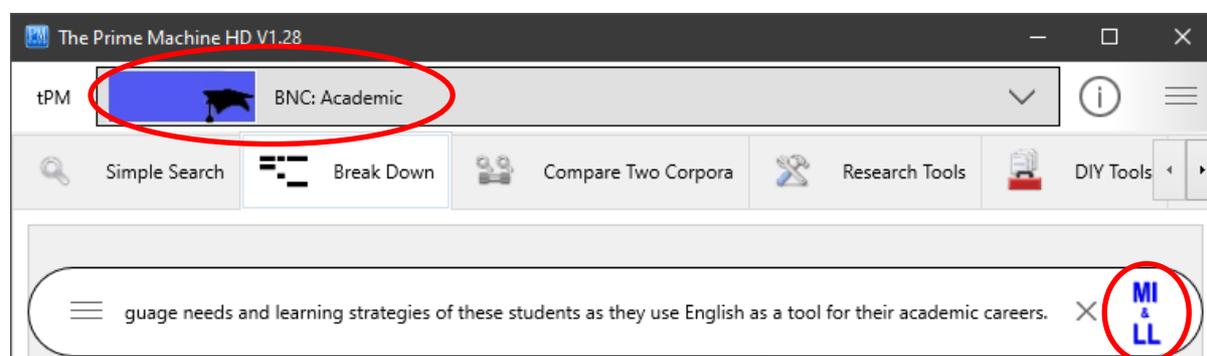


Step 4: Exploring the collocational strength between words in your own sentences

If you want to see which combinations of words in your own sentences have stronger or weaker collocational strengths, you can use the highlighting function in tPM. This will send your combinations of words and look up the collocation scores in the currently selected online corpus. The matches (or missing matches) found in the online corpus will then work as a baseline, showing which words have strong collocational relationships with their neighbours and which do not.

Remember, the collocational strengths of combinations are different in different corpora, so select a suitable corpus as a baseline before you start.

There are two ways to generate these results in tPM. For short strings of words or one sentence, you can use the Break Down tab.



On the Break Down tab, type or paste the sentence into the search box and click or tap the MI & LL button. The combinations of words will be sent to the server and the results will appear as shown below.

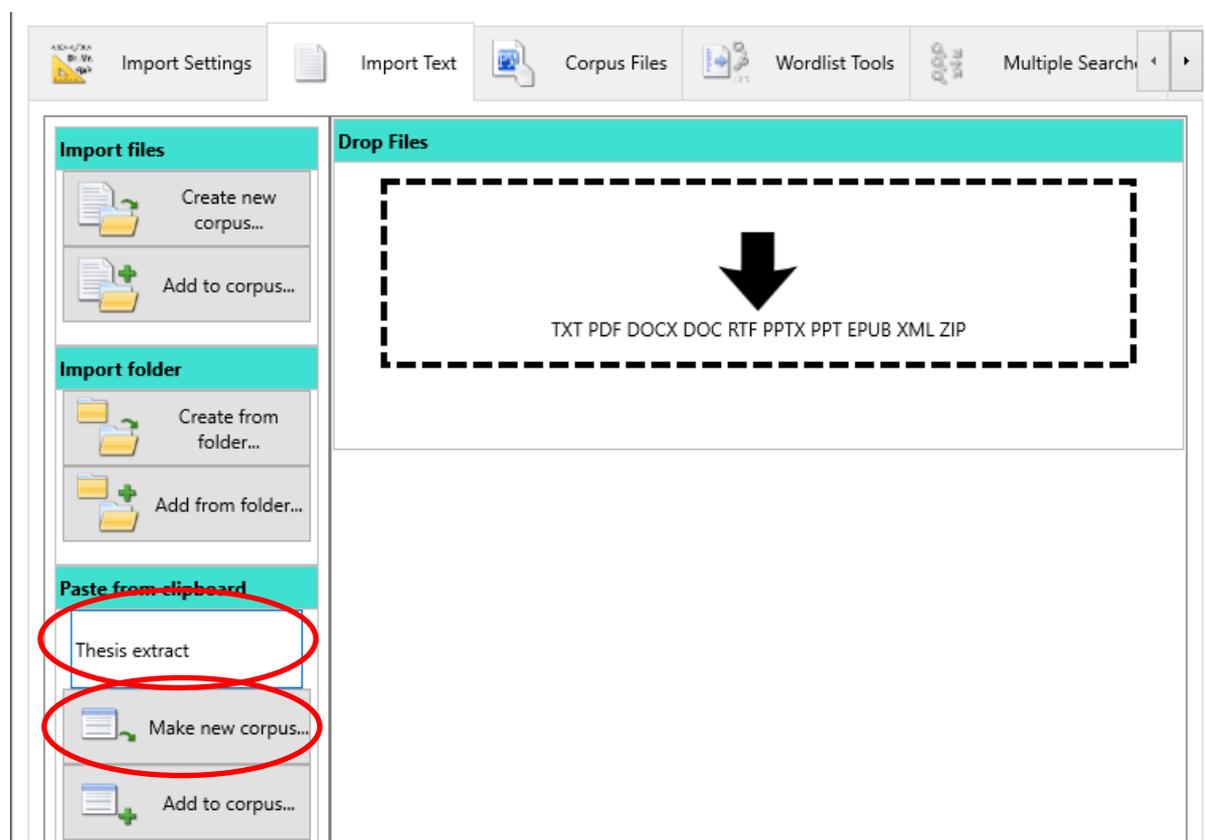
Line	Score	MI	DeltaP	LL	Sentence
1	6.55	100	100	50	as the number of students studying degree subjects through the medium of english grows, it is becoming increasingly important to understand the language needs and learning strategies of these students as they use english as a tool for their academic careers

As well as viewing the results and noting down some combinations to explore later, the numbers in the columns for MI, DeltaP and LL have links, so you can see all the matching collocations and their scores.

In the table below, the LL score of 50 has been clicked, revealing a table of Log-likelihood collocations found in the BNC: Academic corpus which match combinations of words in our sentence.

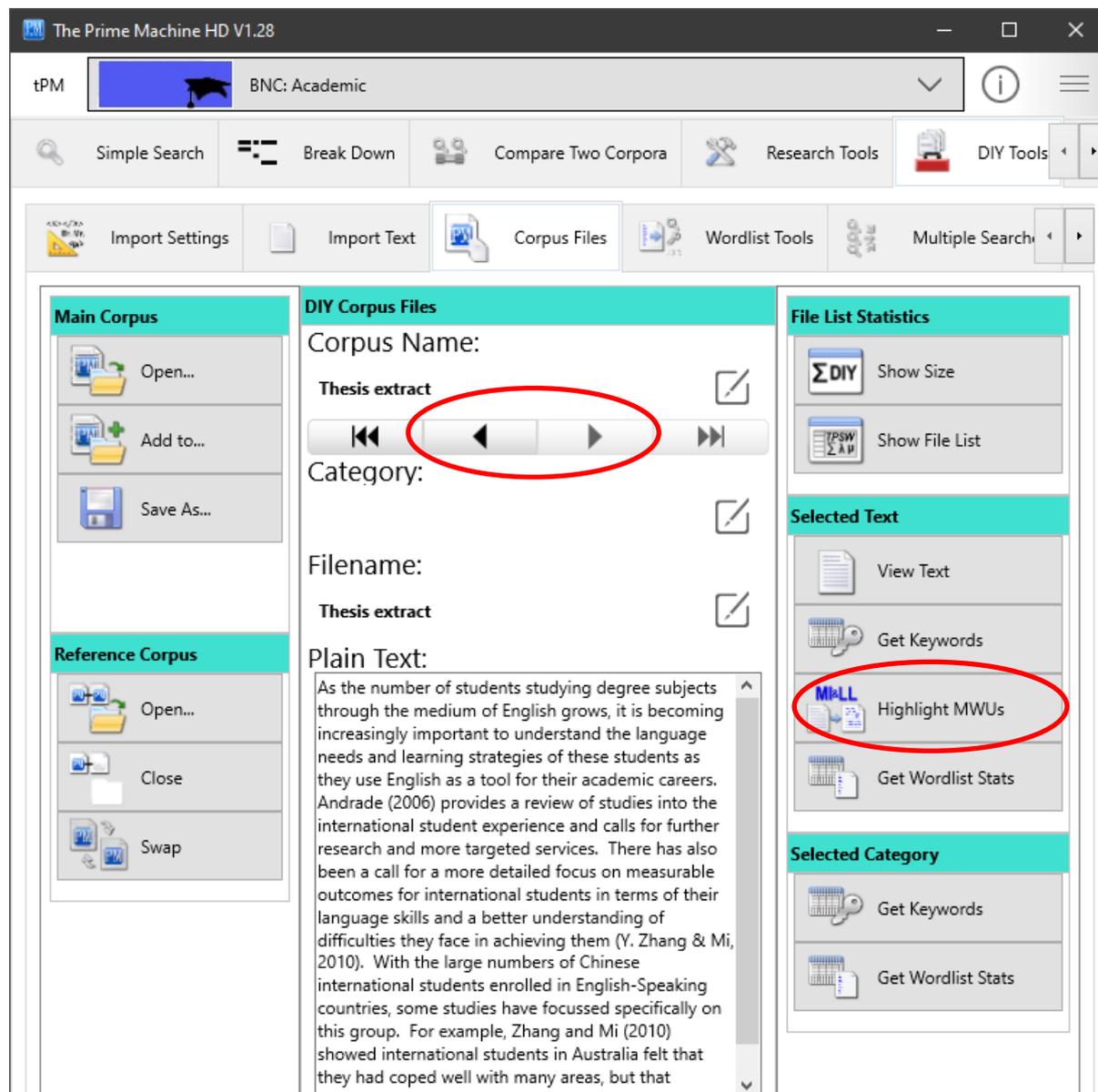
#	Score	LL Collocation	LL
1	10.00	number of	7256.33
2	8.00	number .. students	65.64
3	9.00	through .. medium	96.52
4	10.00	it .. important	1356.37
5	10.00	is .. important	300.95
6	10.00	becoming increasingly	694.99
7	6.00	becoming .. important	39.19
8	10.00	increasingly important	187.75
9	4.00	important .. understand	28.34
10	10.00	to understand	1313.39
11	7.00	language .. learning	45.98
12	6.00	learning strategies	38.87
13	10.00	as .. tool	167.26
14	1.00	a tool	17.05
15	8.00	tool for	70.17
16	7.00	their .. careers	48.56

The other way to get these scores is using one entire text from a DIY corpus. The "Paste from Clipboard" function is a quick way to make a small DIY corpus. You can set the name of the corpus before you paste text using the box above the button.



To generate collocation highlighting results for a DIY corpus text, go to the Corpus Files tab under DIY Tools and find the "Highlight MWUs" button under Selected Text.

If you do want to send an entire text from a DIY corpus, you can navigate the texts on this tab by using the left and right arrows.



When you click the button, combinations of words will be sent to the server.

Note: Using "Highlight MWUs" means chunks of your text will be transmitted over the internet. Efforts are made to make this secure, but you should be mindful of the risks of transmitting these kinds of data.

Line	Score	MI	DeltaP	LL	Sentence
1	6.56	100	100	49	As the number of students studying degree subjects through the medium of English grows , it is becoming increasingly important to understand the language needs and learning strategies of these students as they use English as a tool for their academic careers .
2	5.55	92	92	29	Andrade (2006) provides a review of studies into the international student experience and calls for further research and more targeted services .
3	6.30	100	100	68	There has also been a call for a more detailed focus on measurable outcomes for international students in terms of their language skills and a better understanding of difficulties they face in achieving them (Y .
4	3.50	100	100	29	Zhang & Mi , 2010) .
5	5.86	100	100	50	With the large numbers of Chinese international students enrolled in English - Speaking countries , some studies have focussed specifically on this group .
6	5.37	100	100	45	For example , Zhang and Mi (2010) showed international students in Australia felt that they had coped well with many areas , but that Chinese learners in particular felt they needed more on - going support for academic writing .
7	5.37	98	95	60	In another study , Chen and Duanmu (2010) found that at post - graduate level , the Chinese students who were interviewed struggled more with academic writing than their counterparts and were more likely to employ less active study strategies .
8	5.64	98	98	48	In addition , some researchers have looked at the socialization challenges that face mainland Chinese students in English - medium universities and highlighted the importance of research into understanding how they cope because of the importance of English language learning for the success of these students (Gao , 2010) .

The results break the text into sentences and each sentence has a set of collocation scores. Unlike the Break Down tab, the full stop at the end of each sentence is also included in the scores (so we get slightly different results for the first sentence).

If we look at line 5 in the results above, we can see that the word *focussed* is in grey. This is actually because I have a personal preference for the less usual spelling of this word. It is correct, but *focused* with one *s* seems more common in the BNC: Academic. I can test this theory by changing the spelling and regenerating the results:

Line	Score	MI	DeltaP	LL	Sentence
1	6.65	100	100	50	some studies have focused specifically on this group

But in your own writing, you may be able to find some grey words that intrigue you – some combinations you thought were quite common and didn't need any attention have been coloured grey by this process.

Try to spot some of these and break down the phrases into different combinations, trying to see whether there are synonyms or other wording which will make your expressions seem more natural. Don't take away everything that is creative. But if an expression is grey and it is describing a common idea, you probably should try to change it.

If you are a teacher wanting to analyse your students' work in this way, bear in mind that even in EAP classes the BNC: Academic may be too specialised for the kinds of general EAP topics which often come up in academic writing courses. It may be a good idea to use a more general corpus such as the complete British National Corpus.

tPM is a powerful tool for exploring collocations. Even by following just a few of these clues, you should be able to use your detective and reasoning skills to learn more about the contexts in which writers prefer one word combination over another, and the co-texts that are formed through collocation and grammatical patterns.

For some of the background to these methods and approaches, please see the tPM Help Selected Bibliography available from <https://www.theprimemachine.net/help.html>.



Dr. Stephen Jeaco - 杰大海
www.theprimemachine.net



This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA. If you use this document, please include a link to the tPM Help Selected Bibliography available from <https://www.theprimemachine.net/help.html>.

First published: Thursday, 17 March 2022

Last updated: Thursday, 17 March 2022